Basic words and language evolution*

Feng Wang & William S. -Y. Wang**
Language Engineering Laboratory
City University of Hong Kong

Abstract: In this paper, the basic words introduced by Swadesh (1952, 1955) are used to uncover information about language evolution. In section 2, based on Chen (1996), Swadesh's 200-word list can be split into two sub-groups; the first 100 words, called *high rank*, are more stable and loan-resistant than the second 100, called *low rank*. More retentions tend to be present among the high rank words while borrowed elements tend to occur more frequently among the low rank words. This feature can be used to distinguish retentions and borrowings when the evolution of a language has been blurred by language contact, such as the transmission of Middle Chinese entering tones into Pekinese. In section 3, we argue that Swadesh (1955)'s 100 basic words (high rank) are better for sub-grouping Chinese dialects, compared with Dolgopolsky (1964)'s 15-word list and Yakhontov' 35-word list.

Key words: basic words, rank, Chinese dialect, genetic classification

1 Introduction

Since Swadesh (1952, 1955) put forward 200-word and 100-word lists of basic words[1] for use in glottochronology, different scholars give different basic word lists such as the 15-word of Dolgopolsky (1964) and the 35-word of Yakhontov[2]. Many arguments have focused on whether the change rate of basic words is constant, or whether they are suitable for dating in historical linguistics. Both Cavalli & Wang (1986) and Starostin (1991) have investigated the variations in rate of lexical

* This work was supported in part by grants from the City University of Hong Kong [#9010001], the Hong Kong Research Grants Council [#9040781], and Academia Sinica [Taiwan]. We thank Prof. Tongqiang Xu, who provided the data on some Chinese dialects. We would like to give many thanks to Prof. Baoya Chen and Dr. James W. Minett for kind comments and constructive suggestions. We also thank Prof. Dah'an Ho and Ms. Jinyun Ke for helpful discussion. Thanks are also due to two anonymous reviewers for helpful comments.

** Also University of California, Berkeley.

[1] Some scholars call them kernel words or core words.

[2] The list is quoted from Starostin (1991:59-60).

replacement. In this paper we will explore other usages of basic words. In Section 2, the "ranks" of basic words proposed in Chen (1996) will be taken as a baseline to distinguish borrowing and retention[3]. Chen (1996) splits the 200-word list of Swadesh into two ranks: the *high rank* is just the list of 100 basic words proposed by Swadesh (1955), while the *low rank* is the remainder of the 200 basic words proposed in Swadesh (1952) after excluding the high rank words[4]. Chen (1996) found that words in the high rank tend to be more stable and loan-resistant than those in the low rank, that is, the high rank has more retentions while the low rank tends to be influenced by borrowing more quickly. Based on this finding, we compare the regrouping patterns of Middle Chinese (henceforth MC) entering tones in Pekinese in order to distinguish retention and borrowing in Pekinese. In Section 3, we test three important lists of Swadesh (1955), Yakhontov and Dolgopolsky (1964) in order to find which one is best for sub-grouping Chinese dialects. In the test, a morphological limitation is applied. That is, if all corresponding words of a basic word are compound in all testing languages, that basic word will be discarded. The results show that Swadesh (1955)'s list under the morphological limitation generates the best genetic tree for Chinese dialects.

2 Ranks in basic words

Chen (1996) has proposed that Swadesh's 200 words can be split into two sub-groups: the 100 basic-word list proposed by Swadesh (1955), and the remainder of the Swadesh 200 word-list after removing the first 100 words — Chen calls the two groups *high rank* and *low rank*, respectively. He points out that words in the high rank are more stable and loan-resistant than those in low rank. More retentions of a proto-language will be kept in the high rank, while borrowed elements will be brought into low rank more quickly and easily. Based on this point, Chen (1996) puts forward a method to judge genetic relationships between languages. Words with sound correspondence between languages are called "related words". Chen proposes that languages with genetical relatedness have a greater number of related words in the high rank than in the low rank. On the contrary, the number of related words in the

[3] Starostin (1991) proposed a similar treatment of dividing basic words into two sets according to their replacement rates. See also Comrie (1993) for further discussion.
[4] The remainder is 107-word, which is adjusted to 100-word (c.f. Chen 1996:297).

high rank would be less than that in the low rank if the relationship is just due to language contact. This method was tested on data for languages from several well-established language families, including some Indo-European languages and Chinese dialects. The relationships revealed by this method coincide with well-known judgements. Applying the method to other aspects of language evolution may shed new light on how to determine the origin of linguistic elements.

For a long time, the most complex and intriguing problem in Pekinese has been the discovery of the process by which the MC entering tone (*ru sheng*) regrouped into three tones[5], described succinctly in Chinese as入派三聲. Many scholars have tried to figure out whether the regrouping is the result of internal change or contact-induced change. It is difficult to tell what has been inherited and what has been borrowed in the process.

Bai (1931), Forrest (1950), Hirayama (1960,1990), Stimson (1962), Hsieh (1971), Lin (1992), Ting (1998) and Chen (1999) have interpreted this problem in different ways. Hirayama (1960, 1990) and Stimson (1962) believe that the irregularities in Pekinese are due to the mixture of different dialects. Stimson (1962) devised four "strains" to explain the deposit in Pekinese. Based on lexical diffusion, Hsieh (1971) argues that the varieties are residues of uncompleted sound changes in different periods. Ting (1998) points out, however, this interpretation is too complicated to believe.

When we discuss the evolution of MC tones, we use the traditional Song dynasty classification of initial consonants upon which tonal changes are conditioned with the English notation adopted from Wang (1996):

全清 uu = unvioced, unaspirated

次清 ua = unvoiced, aspirated

全濁 vo = voiced, obstruent

次濁 vs = voiced, sonorant

[5] In MC, the four tones are 平*ping sheng (level tone),* 上*shang sheng (rising tone),* 去*qu sheng (falling tone)* and 入*ru sheng (entering tone).*

Stimson (1962) assigned Pekinese readings to four separate strains, labeled PA, PB, PC and PD, which are transcribed as below:

MC	PA	PB	PC	PD
vo	*yang ping*	*yin ping*	*qu sheng*	*yang ping*
vs	*qu sheng*		*yang ping*	*qu sheng*
uu	*shang sheng*			*yang ping*
ua			*qu sheng*	*qu sheng*

He then stated (1962: 383):

> "A method for determining which of several strains is inherited in a langauage is suggested in an article by Isidore Dyen…, Lg.32.83-7 (1956). This method is quite straightforward: the strain most frequently represented in a short list of basic words is to be considered the inherited strain."

He found 20 readings for the strain PA among the 33 reflexes of MC[6] entering tone syllables in Swadesh's 200-word list. "Thus it is possible to identify the inherited strain as PA."

However, how frequently a strain is respresented in the basic words may be not the point, since it is well know that there is no limit for borrowing. It is possible that a borrowing strain is resprented more frequently in the basic words than the inherited one when the contact is heavy.

We now make use of Chen (1996)'s concept of high rank and low rank to distinguish retention and borrowing. Based on Chen (1996)'s method, if we compare the different regrouping patterns of entering tones in the high rank and the low rank, the pattern that represents the inherited strain will be clarified. If a pattern is borrowed, low rank words would be influenced at first, then reach high rank words. According to this, a pattern that occurs only in high rank words would be inherited from the ancestral language. On the contrary, if a corresponding pattern occurs only in low rank words, it must be borrowed.

The correspondences MC and modern Pekinese in two ranks are listed below.

[6] Stimson called MC Ancient Chinese (AC).

In high rank:

Initial in MC	Tone in Pekinese	Examples
vo	*yang ping*	舌白石
vs	*qu sheng*	月熱葉綠肉
uu & ua	*qu sheng*	血發不
	yin ping	殺吃說一虱膝黑
	shang sheng	骨腳給角

In low rank:

Initial in MC	Tone in Pekinese	Examples
vo	*yang ping*	活薄直
vs	*qu sheng*	獵[7]
uu & ua	*yin ping*	挖擦壓吸濕
	shang sheng	窄雪
	yang ping	結

In both ranks, vo initials cause the entering tones to become *yang ping* and vs initials to become *qu sheng*. The problem is that in Pekinese serveral tones correspond with the entering tone with unvoiced initials in MC without any sound condition. In both ranks, examples corresponding with *yin ping* and *shang sheng* can be found, but some syllables change into *qu sheng* in the high rank basic words, for example, 血發不, while there are no similar examples in the low rank. In low rank, an additional tone corresponding to the entering tone with unvoiced initials in MC is *yang ping*.

If we extend Chen (1996)'s method, corresponding pattern only appearing in the high rank, not in the low rank, is considered to be inherited. In Pekinese, *qu sheng* corresponding to MC entering tone with unvoiced initial is the example. If it is borrowed, there will be some examples to show the same pattern in the low rank since borrowing will influence the low rank words at first. But there is no such trace in Pekinese. On the other side, corresponding pattern only appearing in the low rank, not in the high rank, is very likely to be borrowed. *Yang ping* corresponding to MC entering tone with unvoiced initial is such example.

Therefore, if we recognize that for the entering tone with unvoiced initials in MC the corresponding to *qu sheng* represent the inherited strain, *yin ping* and *shang sheng* are strongly suggested to be from dialects contact with original Pekinese heavily. *Yang ping* variations were borrowed from a different dialect, which interfered Pekinese not so heavily as the two former dialects. The picture of the evolution of the MC entering tone to Pekinese may be summarized in the following table:

Initial in MC	Inherited strain	Borrowed strain (D1)	Borrowed strain (D2)	Borrowed strain (D3)
uu & ua	*qu sheng*	*yin ping*	*shang sheng*	*yang ping*
vs	*qu sheng*	*qu sheng*	*qu sheng*	--
vo	*yang ping*	*yang ping*	*yang ping*	--

Ting (1998) proposed a different hypothesis. According to him, four dialects of Mandarin overlap in Pekinese, resulting in the complex correspondences between MC entering tone and modern tones in Pekinese. The four dialects are listed below:

	Jiaoliao Mandarin 膠遼官話	North Mandarin 北方官話 Shiji Pian 石濟片	Zhongyuan Mandarin 中原官話 Wuhe/Fengyang 五河/鳳陽	Southwest Mandarin 西南官話 Sichuan/Yunnan 四川/雲南
ua & us	*shang sheng*	*yin ping*	*qu sheng*	*yang ping*
vs	*qu sheng*	*qu sheng*	*qu sheng*	*yang ping*
vo	*yang ping*	*yang ping*	*qu sheng*	*yang ping*

Ting (1998)'s proposal may meet some difficulties. If contact of the four dialects in his table results in the pattern of Pekinese, different correspondences to MC entering tone with vo initial, vs initial or unvoiced initial should be observed in Pekinese, because not only for unvoiced initial, but also for vo and vs initial, corresponding patterns of MC entering tone are different in these dialects. In fact, only correspondending patterns for unvoiced initial are variant in modern Pekinese, which is clarified in previous text.

Our analysis based on basic words suggests that in the inherited strain of Pekinese entering tone with unvoiced or vs initials changed into *qu sheng* while entering tone

[7] "拉" is *yin ping* in Pekinese, not *qu sheng*. The reason for this irregularity is unknown.

with vo initials changed into *yang ping*. It then came into contact with two dialects: D1 and D2. D1 and D2 had different tones corresponding to MC entering tone with unvoiced initials, but maintained the same correspondences with other initials. A third dialect, D3, may have interfered with original Pekinese, too, but its influence would not have been so heavy as D1 or D2, because it left no trace among the high rank words of Pekinese.

Our interpretation can be confirmed by other evidence. At first, Guo (1986,1997) and Chen (1999) show that all entering tones with unvoiced initials in the earlier Pekinese changed into *qu sheng* before the Ming dynasty. Chen (1999) points out that according to the statistics and analysis of Guo (1986, 1997), the entering tone with unvoiced initials in MC has variant reflexes during the Ming dynasty: literary pronunciation was *qu sheng* while colloquial pronunciations were *yin ping, shang sheng* or *yang ping*. The literary system of Pekinese was inherited while the colloquial systems were borrowed, due to Pekinese having been the prestige dialect since the Yuan dynasty (1206-1367). All 2,738 characters with unvoiced entering tones have the literary reading — *qu sheng* in the Ming dynasty (1368-1644). This supports our hypothesis about what is the inherited strain of Pekinese.

Secondly, the population movements to today's Pekinese-speaking area support our hyphothesis about the borrowed strains. According to Cao (1997:216-243), the populations from Shandong Province are immigrated to this area at least from the beginning of the Ming dynasty. Lin (1987) ponited out that around the Qing dynasty (1616-1911) many populations from Shandong province are forced to migrate into this area. These migrations make important effects on Pekinese. Based on Qian (2001:21), a classification for dialects in Shandong province is summarized below:

Initial in MC	Eastern dialects	Most of Western dialects
uu & ua	*shang sheng*	*yin ping*
vs	*qu sheng*	*qu sheng*
vo	*yang ping*	*yang ping*

Obviously, the patterns in Eastern dialects and Western dialects in Shandong coincide with D2 and D1, respectively. D3 may be a Southwestern Mandarin, which was brought into the Pekinese-speaking region during the Ming dynasty. A rough map, summarizing this interpretation of population movements, is shown below.

Figure 1: Migrations to Pekinese-speaking area

Two thicker lines indicate the strains from East and West of Shandong; while the thinner line shows the trace of population from Southwestern Mandarin district.

3 Basic words for genetic classification

The 200-basic-word list was originally proposed in Swadesh (1952) for glottochronology. At first, Swadesh believed that these basic words are stable for dating. However, he gradually whittled down the list to 100 words (Swadesh 1955) because some words are found not stable enough. Even for the 100 basic words, many scholars still think this set of words contains a number of problems and doubt its benefits in historical linguistics. So some linguists propose different basic-word lists, such as Dolgopolsky (1964)'s 15-word list and Yakhontov's 35-word list. Since now the lists are mainly used for the same function – genetic classification. The fitness in genetic classification will be an important criterion for basic-word lists.

We devise an algorithm to test the fitness of different basic-word lists. 10 Chinese dialects are selected for testing: Beijing (B), Yingshan (Y), Suzhou (S), Shanghai (H), Shuangfeng (F), Changsha (C), Nanchang (N), Guangzhou (G), Meixian (M), and Xiamen (X). Some genetic relationships between these dialects are well established, for example, B and Y belong to Mandarin; S and H are Wu dialects; C and F are Xiang dialects. Each of the three pairs should be classified into the right sub-group; classifications that fail to do so will be rejected. The three pairs will be taken as the

basic index of fitness of basic words in genetic classification. PHYLIP software will be used to draw genetic tree based on different lists. A similar idea is discussed in Wang (1997). The first step is to determine which words in what Chinese dialects have a common source. The numbers of words from same source between Chinese dialects construct a Similarity Matrix. Since the branches on genetic trees we wish to construct should show distance rather than similarity, it is necessary to transform the Similarity matrix into a Distance matrix. This can be done by deriving each distance, *d*, by taking the negative logarithm of each similarity, *s*; i.e., $d = -\log s$. The distance matrix will be the input for PHYLIP software, and the output will be a genetic tree. Details about the application are discussed in Saitou & Nei (1987) and Wang (1997).

Before testing, we would like to restate a morphological criterion for applying the basic-word list into individual languages. From morphological point of view, it is well know that root-words are more basic than compound words. For instance, a language may refer to "moon" as "eye of the night". Obviously "eye" and "night" are more basic than "moon" in this language since "moon" is then a compound word, not an irreducible form. So we can discard it from the basic-word list. Sapir (1916:434) pointed out:

> "One of the most useful principles for the determination of the age of a word is a consideration of its form; that is, whether it can be analysed into simpler elements, its significance being made up of the sum of these, or is a simple irreducible term. In the former case we suspect, generally speaking, a secondary or relative late formation, in the latter considerable antiquity."

The strategy is to exclude those words that fail to satisfy the morphological criterion according to different individual languages. We then obtain a list of "relative basic words" for different languages. In Chinese, some words are compound in all dialects. For example, "bark" corresponds to "樹 tree 皮 skin" in all Chinese dialects, which suggests that it is unsuitable to be taken as a basic word in Chinese dialects.

Therefore, 5 basic-word lists will be tested: 1) Dolgopolsky (1964) has investigated the stability of 15 meanings among the language families of Northern Eurasia. His list

is: *first person marker, two, second person marker, who/what, tongue, name, eye, heart, tooth, verbal negation, finger-nail/toe-nail, louse, tear (noun), water, and dead*. "*who/what*" will correspond to two words in each Chinese dialect. So we get the 16-word list. 2) Two words- *tear* and *nail* in the 16-word list are compound words in all Chinese dialects, which will be discarded according to the morphological criterion. We therefore obtain a 14-word list. 3) Yakhontov has proposed the 35 most stable meanings: *blood, bone, die, dog, ear, egg, eye, fire, fish, full, give, hand, horn, I, know, louse, moon, name, new, nose, one, salt, stone, sun, tail, this, thou, tongue, tooth, two, water, what, who, wind, year*. For Chinese dialects, we need not discard any words according to the morphological criterion. 4) Swadesh's 100 basic-word is the fourth list. 5) According to the morphological criterion, five words, *hair, swim, woman, man, bark,* in the forth list will be excluded. Then the rest 95 words will construct the fifth list.

In fact, two features are tested in the schedule above. One is the fitness of the basic-word lists proposed by Dolgopolsky, Yakhontov and Swadesh, respectively. The other is the morphological criterion.

We will use the form like ((A,B)C) to represent the genetic relationship between language A, B and C. ((A,B)C) means that A and B are closer to each other than either is to C. Based on Dolgopolsky's list and Yakhontov's list, results for genetic relationship by PHYLIP are listed as below:

List	Result
14-word	((S,H)((M,X)((N(C,F,Y))(G,B))))
16-word	((((G(((C,Y)F)(B,N)))(S,H))(M,X))
35-word	(((B,G)(M,X))((S,H)((C,F)(N,Y))))

For each of the three lists, B and Y are separated, which cannot be accepted. The results show that the 14-word, 16-word and 35-word lists do not satisfy our expectation of the sub-grouping (B,Y). This suggests that Yakhontov's 35-word list and Dolgopolsky's list are not suitable for sub-grouping Chinese dialects.

For Swadesh' 100 words and the 95 basic words, the three pairs of dialects form the subgroups as expected. In order to compare them, an additional parameter, stability, is introduced to measure results. In this test, the standard representatives (B, N, C, S, G, M, X) of Seven Major Chinese dialects (Mandarin, Gan, Xiang, Wu, Yue, Hakka, Min) will be used as fixed items, and the other three dialects whose genetic positions are well-known will be taken as optional items, such as F in Xiang dialect, H in Wu dialect and Y in Mandarin[8]. We assume the adding of optional items may not modify the topology of the fix items that much if the right tree will be generated based on a certain list. Based on each list, we will obtain a group of topologies for the fixed items by adding the optional items to fixed items one by one. The distances between the topologies in each group will be calculated as the index of stability of topology, and the smallest one may indicate the best fitness of a list. The results are listed below:

1) 100-word list:

Optional item	Result
+F	(((B(S(N(C,F))))G)(M,X))
+H	(((B(S(C,N)))G) (M,X))
+Y	(((((B,Y)C)(S,N))G) (M,X))
+F,+H	((((B(S,H))(N(C,F)))G) (M,X))
+F,+Y	(((((B,Y)(C,F))(S,N))G) (M,X))
+Y,+H	((((B,Y) C) ((S,H)N) G) (M,X))
+F,+H,+Y	(((((B,Y)(C,F))(N(S,H)))G) (M,X))

2) 95-word list:

Optional item	Result
+F	(((B((S,N)(C,F)))G) (M,X))
+H	(((B((S,H)(N,C)))G) (M,X))
+Y	(((((B,Y)C)(S,N))G) (M,X))
+F,+H	((((B(S,H))(N(C,F)))G) (M,X))
+F,+Y	(((((B,Y)(C,F))(S,N))G) (M,X))
+Y,+H	(((((B,Y)C)((S,H)N))G) (M,X))

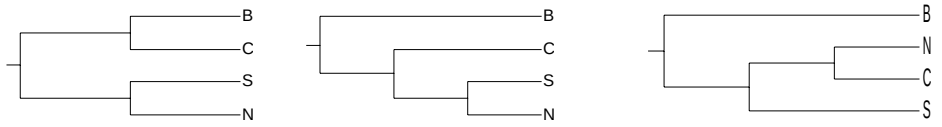
[8] We thank all the experts who provided the data of Chinese dialects. They are Prof. Mei Fang, Prof. Xiaofan Li and his student, Ms. Yan Xiong, Prof. Eric Zee, Prof. Yun Mai, Prof. Xiuhong Yan, Prof. Baokui Ye and his student, Ms. Ruiyuan Xu.

+F,+Y,+H	(((((B,Y)(C,F))((S,H)N))G) (M,X))

Comparing topologies of the seven fixed dialects, the positions of G, M and X are always unchanged. That means, they do not afford any diagnostic information for topologies, then they will be ignored when differences between topologies are calculated. For 100-basic-word list, we got three different types: 1. (((N,C)S)B); 2. ((B,C)(S,N))); 3.((B,S)(N,C)). For 95-word list, we got four types: 1. (((S,N)C)B); 2. (B(S(N,C))); 3.((B,S)(N,C)); 4.((B,C)(S,N)). The minimum movements from one topology to another will measure the topological distance. For example:

Dist { ((B,C)(S,N)) , (((N,C)S)B) }:

((B,C)(S,N)) ---> (B(C (S,N))) ---> (B (S (C,N)))=(((N,C) S) B)



2 movements => distance = 2

According to this algorithm, the sum of distances between all topologies based on 100 words is 24, while the sum for 95 words is 22. That is to say, the 95 words result in more stable topology for the seven Major Chinese dialects. It suggests that the 95-word list is more suitable for Chinese dialects.

In order to test the efficacy of the deletion according to the morphological criterion, we deleted five words randomly and repeat the same procedure as above to classify Chinese dialects. Three experiments have been done. The three groups of five words are: 1. *skin, knee, ash, stone, I*; 2.*nose, smoke, walk, seed, dog*; 3. *bird, grease, star, all, cloud*. With the three kinds of random deletion, the sums of distances between the typologies are 24, 24 and 34, respectively. All numbers are larger than the sum 22 which is based on the morphological criterion. The tests show that Swadesh's 100 basic words adjusted by the morphological criterion are of the best fitness in genetic classification for Chinese dialects.

4 Discussion

Basic-words are an important window to language evolution. A group of Chinese dialects was taken as a case to explore usages of basic words. The distinction between high/low ranks (Chen 1996) in basic words is used to distinguish retention and borrowing, and high-rank words show the fitness in application to genetic classification on Chinese dialects. However, each word may have a particular replacement rate, maintained on average in different language groups (c. f. Cavalli & Wang 1986). There is still no convincing answer to the question why the boundary of high / low words should be set as in Chen (1996). How to sort basic words according to their change rates still needs much more substantive research. It may be necessary to look for factors causing the variations of replacements of words. Under 'well-controlled' conditions, it would be promising to answer whether there is a universal boundary between high/low rank basic words in world languages and how to determine it. Whatever the answer is, the exploration will bring more information on language evolution.

Appendix A- High rank words (the word with entering tone is marked in black font)

1 I 我	2you 你	3we 我們	4this 這	5that 那
6who 誰	7what 什麼	8not **不**	9all 全部	10many 多
11one **一**	12two 二	13big 大	14long 長	15small 小
16women 女人	17man 男人	18person 人	19fish 魚	20bird 鳥
21dog 狗	22louse **蝨**子	23tree 樹	24seed 種子	25leaf **葉**子
26root 根	27bark 樹皮	28skin 皮膚	29flesh **肉**	30blood **血**
31bone **骨**頭	32grease 脂肪	33egg 雞蛋	34horn **角**	35tail 尾巴
36feather 羽毛	37hair 頭**髮**	38head 頭	39ear 耳朵	40eye 眼睛
41nose 鼻子	42mouth 嘴	43tooth 牙齒	44tongue **舌**頭	45claw 爪子
46foot **腳**	47knee **膝**蓋	48hand 手	49belly 肚子	50neck 脖子
51breasts 乳房	52heart 心臟	53liver 肝	54drink 喝	55eat **吃**
56bite 咬	57see 看見	58hear 聽到	59know 知道	60sleep 睡
61die 死	62kill **殺**	63swim 游水	64fly 飛	65walk 走
66come 來	67lie 躺	68sit 坐	69stand 站	70give **給**
71say **說**	72sun 太陽	73moon **月**亮	74star 星星	75water 水
76rain 雨	77stone **石**頭	78sand 沙子	79earth 土地	80cloud 雲
81smoke 煙	82fire 火	83ash 灰	84burn 燒	85path 路
86mountain 山	87red 紅	88green **綠**	89yellow 黃	90white **白**
91black **黑**	92night 晚上	93hot **熱**	94cold 冷	95full 滿
96new 新	97good 好	98round 圓	99dry 幹	100name 名字

Appendix B- Low rank words

1and 和	2animal 動物	3back 背	4bad 壞	5because 因爲
6blow 吹	7breathe 呼**吸**	8child 孩子	9count 數	10cut 砍
11day 天	12dig **挖**	13dirty 髒	14dull 呆、笨	15dust 塵土
16fall 掉	17far 遠	18father 父親	19fear 怕	20few 少
21fight 打架	22five 五	23float 漂浮	24flow 流	25flower 花
26fog 霧	27four 四	28freeze **結**冰	29fruit 水果	30grass 草
31guts 腸子	32he 他	33here 這裏	34hit 打	35hold/take 拿
36how 怎麼	37hunt 打**獵**	38husband 丈夫	39ice 冰	40if 如果
41in 在	42lake 湖	43laugh 笑	44leftside 左邊	45leg 腿
46live(alive) **活的**	47mother 母親	48narrow **窄**	49near 近	50old 老
51play 玩	52pull 拉	53push 推	54rightside 右邊	55correct 對
56river 江	57rope 繩子	58rotten 腐爛	59rub **擦**	60salt 鹽
61scratch 抓	62sea 海	63sew 縫	64sharp 尖	65short 短
66sing 唱	67sky 天空	68smell 聞	69smooth 平	70snake 蛇
71snow **雪**	72spit 吐	73split 撕裂	74squeeze **壓**	75stab 刺
76stick 棍子	77straight **直**	78suck 吮	79swell 腫	80there 那兒
81they 他們	82thick 厚	83thin **薄**	84think 想	85three 三
86throw 扔	87tie 捆	88turn 轉	89vomit 嘔吐	90wash 洗
91wet **濕**	92where 哪里	93wide 寬	94wife 妻子	95wind 風
96wing 翅膀	97heavy 重	98woods 森林	99worm 蟲	100year 年

Appendix C- 100 Basic words in the Chinese dialects:

Notation: 1. Some special Chinese characters or some words without character representations will be represented by Capital alphabets. 2. Only the reflexes of a word are the same in two dialects, they will be counted as the same state. For the word "*black*", 黑/青 in F and 黑 in H is recognized as two different states.

	G	S	M	N	Y	B	C	F	H	X
all	鹹	通	完	都	下	整	下	下	全	全
ash	灰	灰	灰	灰	灰	灰	灰	灰	灰	灰
bark	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮
belly	肚	肚	肚	肚	肚	肚	肚	肚	肚	腹肚
big	大	大	大	大	大	大	大	大	A	大
bird	雀/鳥	鳥	鳥	鳥	雀	鳥	鳥	鳥	鳥	A
bite	咬	咬	咬/齧	咬	咬	咬	咬	咬	咬	咬
black	黑	黑	烏	烏/青/黑	黑	黑	黑/青	黑/青	黑	烏
blood	血	血	血	血	血	血	血	血	血	血
bone	骨	骨	骨	骨	骨	骨	骨	骨	骨	骨
breasts	A	奶	奶	奶	媽	咂/媽	奶	奶	奶	奶
burn	燒	燒	燒	燒	燒	燒/著	燒	燒	燒	燒/熱
claw	爪	腳爪	腳爪	爪/腳爪	爪	爪	爪	爪	腳爪	爪
cloud	雲	雲	雲	雲	雲	雲	雲	雲	雲	雲
cold	凍/冷	冷	冷	冷	冷	冷	冷/清	冷/清	冷	冷/寒
come	來/嚟	來	來	來	來	來	來	來	來	來
die	死	死	死	死	死	死	死	死	死	死
dog	狗	狗	狗	狗	狗	狗	狗	狗	狗	狗
drink	飲	吃	食	吃	喝	喝	吃	吃/呵	吃	啉/啜
dry	幹	幹/燥	A	幹	幹	幹	幹	幹/A	幹	焦
ear	耳	耳	耳	耳	耳	耳	耳	耳	耳	耳
earth	地	地	地	地	地	地	地	地	地	地
eat	吃	吃	食	吃	吃	吃	吃	吃	吃	食
egg	春	蛋	卵	蛋	蛋	雞子/蛋	蛋	蛋	蛋	卵
eye	眼	眼	目	眼	眼	眼	眼	眼	眼	目瞅
feather	毛	羽毛	毛	毛	毛	毛	毛	毛	毛	毛
fire	火	火	火	火	火	火	火	火	火	火
fish	魚	魚	魚	魚	魚	魚	魚	魚	魚	魚
flesh	肉	肉	肉	肉	肉	肉	肉	肉	肉	A

fly	飛	飛	飛	飛	飛	飛	飛	飛	飛	飛
foot	腳	腳	腳	腳	腳	腳	腳	腳	腳	骹
full	滿	滿	滿	滿	滿	滿	滿	滿	滿	A
give	畀	撥	分	把/給	把	給	把	B	撥	A
good	好/A	好/美/贊	好/A	好	好	好/強	好	好	好/靈光	好
grease	油/膏	油	油	油	油/膘	大油/葷油	油	油	油	油
green	綠	綠	青	綠	綠	綠	綠	綠	綠	青
hair	頭髮	頭髮	頭顱毛	頭髮	頭髮	頭髮	頭髮	頭髮	頭髮	頭毛
hand	手	手	手	手	手	手	手	手	手	手
head	頭	頭	頭顱	頭	腦	腦	腦	腦	頭	頭
hear	聽	聽	聽	聽	聽	聽	聽	聽	聽	聽
heart	心	心	心	心	心	心	心	心	心	心
horn	角	角	角	角	角	角	角	角	角	角
hot	熱/慶	熱	燒	熱	熱	熱	熱	熱	熱	熱/燒
i	我	我	我	我	我	我	我	我/卬	我	我
kill	殺	殺	殺	殺	殺	殺	殺	殺	殺	殺
knee	膝	膝饅頭/腳饅頭	膝	舌頭蓋	膝	膊楞蓋兒	膝	膝	腳饅頭	骹頭 A
know	知	曉	知	曉	曉	知	曉	曉	曉	知
leaf	葉	葉	葉	葉	葉	葉	葉	葉	葉	箬
lie	睏	睏/A	眠	睏	睡	躺	困	睏	睏	倒
liver	肝	肝	肝	肝	肝	肝	肝	肝	肝	肝
long	長	長	長	長	長	長	長	長	長	長
louse	虱	虱	虱	虱	虱	虱	虱	虱	虱	虱
man	男人/佬	男/男子客	男子人	男個	男的/男將	男的/爺們	男的/男人家	男人家	男人	A 夫/A 夫儂
many	多	多/交關	多	多	多	多	多	多	多	A
moon	月	月	月	月	月	月	月	月	月	月
mountain	山	山	山	山	山	山	山	山	山	山
mouth	嘴/口	嘴	啜	嘴	嘰	嘴	嘴	嘴	嘴	喙
name	名	名	名	名	名	名	名	名	名	名
neck	頸	頸/頭根	頸	頸	頸	脖	頸	頸	頸	頷管
new	新	新	新	新	新	新	新	新	新	新
night	夜晚/晚黑	夜	夜	夜	黑	夜/黑	夜	夜	夜下昏時/冥時/暗時	
nose	鼻	鼻	鼻	鼻	鼻	鼻	鼻	鼻	鼻	鼻
not	唔/冇	勿	唔	不	不	不	不	不	勿	勿會/無
one	一	一	一	一	一	一	一	一	一	一
path	路	路	路	路	路	道/路	路	路	路	路

person	人	人	人	人	人	人	人	人	人	人
rain	雨	雨	雨	雨	雨	雨	雨	雨	雨	雨
red	紅	紅	紅/赤	紅	紅	紅	紅	紅	紅	紅
root	根/蓾	根	根	根	根	根	根/篦	根/莞	根	根
round	圓	圓	圓	圓	團	圓	圓	欒	圓	圓
sand	沙	沙	沙	沙	沙	沙	沙	沙	沙	沙
say	講/話	說/講	講/話	話	說	說	講	話/講	講	講
see	睇	看/望	看	看/望/妻*	看	看/瞧/瞅/A	看	看/相	看	看
seed	種	種	種	籽	種	種	種	種	種	種/籽
sit	坐	坐	坐	坐	坐	坐	坐	坐	坐	坐
skin	皮	皮膚	皮	皮	皮	皮	皮	皮	皮	皮
sleep	睏	睏	睡	睏	睡	睡	困	睏	睏	睏
small	細	小	細	小/細	小	小	細/小	細	小	細/小
smoke	煙	煙	煙	煙	煙	煙	煙	煙	煙	薰
stand	徛	立	徛	徛/站	站	站/戳	站/企	徛	立	徛
star	星	星	星	星	星	星	星	星	星	星
stone	石	石	石	石	石	石	石	石	石	石
sun	頭	太陽	日	日	日	太陽	太陽/日	日/太陽	太陽	日
swim	游水	游水	泅水/洗身	玩水/洗澡	抹汗	凫水/游泳	洗冷水澡/游泳	洗冷水澡	游泳	泅水
tail	尾	尾	尾	尾	尾	尾	尾	尾	尾	尾
that	個個	歸個/A 個	B 個	C	那	那個	那個/那只	喏只	伊	許
this	爾個	該個/哀個/A 個	B 個	C	D	這個	F 個	咯只	G	即
tongue	脷	舌	舌/利	舌	舌	舌	舌	舌	舌頭	舌
tooth	牙齒	牙齒	牙齒	牙齒	牙齒	牙	牙齒	牙齒	牙	喙齒
tree	樹	樹	樹	樹	樹	樹	樹	樹	樹	樹
two	二	二	二	二	二	二	二	二	兩/二	二
walk	行	走/跑	行	走	走	走	走	行	走	行
water	水	水	水	水	水	水	水	水	水	水
we	我哋	佢	我兜人	我裏/我們	我們	我們	我們	我哩/卬哩	我 A/阿拉	阮
what	乜	啥	A	什	麼	什麼	麼	麼	啥 B 事體	什物
white	白	白	白	白	白	白	白	白	白	白
who	邊個	啥人/A 格	瞞人	哪個	哪個/啥個	誰	哪個	哪個	啥人	啥人/AB
woman	女人/婆嬤	女 A	婦人家	女個	女的/女將	女的/娘們	女的/女人家/堂客	女人家/堂客們	女人	查某/女的
yellow	黃	黃	黃	黃	黃	黃	黃	黃	黃	黃
you	爾	汝	爾	爾	爾	爾	爾	爾	儂	爾

Reference

Bai, Dizhou (白滌洲) 1931. Beiyin rusheng yinbian kao [北音入聲音變考]. Beijing: Beijing nvzi shifan daxue *Xueshuyuekan[*北京女子師範大學《學術月刊》*]* 2. 2.

Cao, Shuji (曹樹基) 1997. Zhongguo yiminshi [中國移民史]. Fuzhou: Fujian renmin chubanshe[福建人民出版社] v.5.

Cavalli-Sforza, L. L. & Wang, W. S-Y. 1986. Spatial Distance and Lexical Replacement. *Language* 62:38-55.

Chen,Baoya (陳保亞) 1996. Lun yuyanjiechu yu yuyanlianmeng [論語言接觸與語言聯盟]. Beijing: Yuwen chubanshe [語文出版社].

Chen,Baoya (陳保亞) 1999. 20 Shiji zhongguo yuyanxue fangfalun [20 世紀中國語言學方法論]. Ji'nan: Shandong jiaoyu chubanshe[山東教育出版社].

Comrie, B. 1993. Review of Starostin (1991). *Language* 69.3: 828-832.

Dolgopolsky, A. B. 1964. A Probabilistic Hypothesis Concerning the Oldest Relationships Among the Language Families in Northern Eurasia, Original Russian translated into English and published in V. V. Shevoroshkin and T. L. Markey (eds.): 1986, *Typology, Relationship and Time*, Ann Arbor: Karoma Publishers.

Dyen, Isidore 1956. The Ngaju-Dayak 'Old Speech Stratum'. *Language* 32.1: 83-7

Forrest, R. A. D. 1950. The Ju-sheng Tone in Pekingese. *Bulletin of the School of Oriental and African Studies* 13.2:443-447.

Guo, Li (郭力) 1986. *Chongding shimawen'gong dengyuntujing* yanjiu [《重訂司馬溫公等韻圖經》研究]. Beijing:Beijingdaxue zhongwenxi shuoshi xueweilunwen 北京大學中文系碩士學位論文.

Gou,Li(郭力) 1997. Gu qingruzi zai *Hebingzixuejiyun* zhong de guidiao[古清入字在《合併字學集韻》中的歸調]. *Yuyanxueluncong[*語言學論叢*]* 19.

Hirayama Hisao(平山久雄) 1960. Zhonggurusheng zai Beijinghuali de duiyingtongze[中古入声在北京话里的对应通则], *Riben Zhongguoxuehuibao [日本中國學會報]* 12:139-156.

Hirayama Hisao(平山久雄) 1990. Zhongguhanyu de qingruzi zai Beijinghuali de duiyingguilv [中古漢語的清入字在北京話裏的對應規律]. *Beijingdaxuexuebao[北京大學學報]* 5：72-79.

Hsieh, Hsin-i 1971. The Development of Middle Chinese Entering Tone in Pekinese. California:University of California at Berkeley Dissertation.

Lin, Tao(林燾) 1987. Beijing guanhua suyuan[北京官話溯源]. *Zhongguoyuwen*[中國語文] 1987.3.

Lin, Tao(林燾) 1992. "Rupaisansheng" bushi[“入派三聲”補釋].*Yuyanxue luncong*[語言學論叢] 17.

Qian, Zengyi (錢曾怡) 2001. Shandong fangyan yanjiu [山東方言研究]. Ji'nan: Qilu shushe[齊魯書社].

Saitou, N. & M. Nei. 1987. The Neighbor-joining Method : a New Method of Reconstruciting Phylogenetic Trees. *Moleclar Biology and Evolution* 4.406-425.

Starostin. S. 1991. Altajskaja Problema i Proisxoz📖denie Japonskogo Jazyka [The Altaic problem and the origin of the Japanese language]. Moscow: Nauka. 59-60.

Sapir.E. 1916. Time Perspective in Aboriginal American Culture: a Study in Method, Canada, Department of Mines, Geological Survey, *Memoir 90 Anthropological Series* No.13. Also in *Selected writtings of Edward Sapir*, Univeristy of California Press,1949.

Stimson, H.M. 1962. Ancient Chinese -p, -t, -k Endings in the Peking Dialect, *Language* 39.4:376-384.

Swadesh. M. 1952. Lexico-statistic dating of prehistoric ethnic contacts, *Proceedings of the American philosophical society* 96.4.

Swadesh. M 1955. Time Depths of American Linguistic Groupings. *American Anthropologist* 56.

Ting, Pang-Hsin (丁邦新) 1998. Hanyu fangyan jiechu de jige leixing [漢語方言接觸的幾個類型].*Yuyanxu luncong[語言學論叢]* 20.

Wang, William S.-Y. 1996. Linguistic Diversity and Language Relationships. *New Horizons in Chinese Linguistics*, ed. by C.-T. James Huang & Y.-H. Audrey Li. Dordrecht; Boston: Kluwer Academic Publishers.

Wang, William S.-Y. 1997. A Quantitative Study of Zhuang-Dong Languages. *In Memory of Mantaro J. Hashimoto,* ed. by Anne O. Yue & Mitsuaki Endo. Tokyo: Uchiyama: 81-96.

基本詞彙與語言演變

汪鋒 王士元

語言工程實驗室

香港城市大學

提要： 本文利用 Swadesh (1952, 1955)提出的基本詞彙表來探討語言演變中的一些問題。根據陳保亞（1996）劃分詞階的方法，Swadesh 的 200 詞可一分爲二： 第 100 詞稱爲高階詞；餘下的 100 詞稱爲低階詞。高階詞比低階詞更穩定，更難以借用。因此，高階詞的同源保留率比低階詞高，而低階詞的借用率比高階詞高。本文第 2 節根據這一規律來厘清北京話入聲字中的早期遺存和晚近的借用成分， 並藉以說明這一規律在語言演變研究中的重要作用。本文第 3 節比較了幾種影響較大的基本詞彙表－Dolgopolsky (1964) 15 詞，Yakhontov 35 詞和 Swadesh (1955) 100 詞，觀察它們在漢語方言分區中的功效，發現根據 Swadesh (1955) 100 詞得出的結果更可信。同時，複合詞應該從基本詞中剔除的看法得到了驗證。

關鍵字：基本詞彙, 階, 漢語方言, 系屬分類